Associations between characteristics of students (ABCS) 2021

Annexes

Reference OfS 2021.46

Enquiries to Paula Duffin at official.statistics@officeforstudents.org.uk

Publication date 13 October 2021

Contents

Annex A: Modelling methodology

1. The ABCS modelling methodology uses a binary logistic regression model with all characteristics forced in as main effects and every two-way interaction between categories tested using a stepwise regression method with entry and stay criteria of $\alpha=0.05$. This decides which of the two-way interactions are included in the final model.

2. We have carried out some sensitivity analysis to determine if it was viable to run the models with fewer years of data rather than the full five-years' worth. We ran all three models with three- and four-years' worth of data and compared the resulting numbers of groups in each quintile with the numbers of groups in each quintile having run with the five years of data. The majority of groups remained in the same quintile whichever time series was used and the small proportions that had changed quintile tended to move up or down by only one quintile, with very few changing by more than this. By this method we were also able to see the number of groups that were placed into quintiles using the full five years of data that simply didn't exist when using only three of four years of data, as no students had these combinations of characteristics across these years. The numbers of these groups were significant, especially in the case of the FT continuation model, although absolute numbers of students in these groups were small. For this reason, we have chosen to continue running our models with five years of data in order to maximise the number of unique groupings that we derive quintiles for.

3. In the same way that we carried out the sensitivity analysis, we also compared the results from our three models with those from the last update in November 2020. Despite some minor changes to the methodology and rolling forward the data a year, the quintiles that are derived in this update are broadly the same as before. For access, 90.9 per cent of students remain in the same quintile as previously. A further 1.5 per cent hold characteristic combinations that had not been seen in previous years, and so comparison with previous quintiles wasn't possible. For FT continuation, 75.1 per cent of students are in the same quintile as previously. However, this is largely accounted for by the fact that 11.2 per cent of students held combinations of characteristics that had not been previously seen. All of the new student groups (where there are combinations of characteristics not previously seen) are very small, typically only including one or two students. PT continuation sits between the other two measures, with 85.7 per cent of students remaining in the same quintile as before.

4. Where student groups are in a different quintile in this latest update, the vast majority have only moved up or down by one quintile. This assures us that the minor changes to the methodology and rolling the time series forward by a year for each measure has not resulted in drastic changes to how student groups are split across the quintiles.

Annex B: Characteristic definitions

1. The characteristics used are all based on variables available on the Department for Education's National Pupil Database (NPD), the Education and Skills Funding Agency's (ESFA's) individualised learner record (ILR) or the Higher Education Statistics Authority's (HESA's) student record or alternative provider (AP) student record.

2. Income Deprivation Affecting Children Index (IDACI) and the Index of Multiple Deprivation (IMD) quintiles are obtained from the Ministry of Housing, Communities and Local Government (MHCLG) and are added to the NPD, ILR, HESA student and AP student records.[1]

Age

3. Age, as used in ABCS FT and PT continuation measures, is defined as the age of the student on the 31 August in the year in which they started their course. This is calculated using their birth date as reported on the ILR or HESA student or student alternative records. Age is not used as a characteristic in the access measure as the base population is already restricted to 18- and 19-year-olds.

Care experience

4. Care experience describes whether a student has ever spent time in the care of a local authority in England or Wales, or Health and Social Care Trust in Northern Ireland, or who has self-declared as being in care for three months or more. The data is collected on the HESA student record only. The definition and reporting of care experience is complex. For details of the data used, see www.hesa.ac.uk/collection/c19051/a/careleaver. Care experience is only used as a characteristic in the FT continuation measure.

Disability

5. Disability information is only used from the ILR and HESA student and AP student records. Disability is self-reported by students. Details of the disability types used can be found under the IPDISABLETYPE section in OfS's 2021 Core Algorithms document.[2] Disability is used as a characteristic in the FT and PT continuation measures.

Ethnicity

6. Ethnicity is used as a characteristic in all three ABCS measures. Ethnicity information is taken from the NPD for the access measure and from the ILR and HESA student and AP student records for the continuation measures. For ABCS access we use the groupings as they are collected in the NPD, with the exception of Gypsy/Roma and Traveller of Irish

[1] See www.gov.uk/government/statistics/english-indices-of-deprivation-2019

[2] See www.officeforstudents.org.uk/data-and-analysis/institutional-performance-measures/technical-documentation/

heritage, which are combined to make a single group, since the groups are very small when separated.

7. For continuation, we use the detailed groupings that the OfS reports by. More detail can be found under the IPETHNICDETAIL section OfS's 2021 Core Algorithms document.[3]

FSM eligibility

8. FSM eligibility indicates whether the student was ever recorded as being eligible to receive free school meals in the six years prior to the March census date in their final year of key stage four (year 11). This characteristic is used in the access and FT continuation measures.

9. The data on free school meals eligibility is produced by the Department for Education (DfE) as part of the National Pupil Database (NPD) and was linked onto data in the ILR and HESA student and student alternative records.[4] For this analysis, the data is restricted to only those who attended a state-funded mainstream school and were under 21 at the start of their higher education studies.

IDACI

10. The Income Deprivation Affecting Children Index (IDACI) is used in all three ABCS measures. It measures the proportion of children under the age of 16 in low-income households for a particular area in England.[5] It is calculated at lower-layer super output area (LSOA) level and is a supplementary measure to the Index of Multiple Deprivation (IMD), and therefore has been constructed for the Ministry of Housing, Communities and Local Government (MHCLG) by Oxford Consultants for Social Inclusion (OCSI) and deprivation.org. In this analysis, we use IDACI quintiles, where the most deprived areas are in quintile 1 and the least deprived are in quintile 5. For pupils or students domiciled outside of England, the value is set to 'Unknown or N/A', and data is not presented in the interactive tools.

IMD

11. The Index of Multiple Deprivation (IMD) is a measure of levels of deprivation for a small area within England and is used in all three ABCS measures.[6] It is calculated at lower-layer super output area (LSOA) level and uses several different measures to determine levels of deprivation. They have been constructed for the Ministry of Housing, Communities and Local Government (MHCLG) by Oxford Consultants for Social Inclusion (OCSI) and deprivation.org. In our analysis, we use IMD quintiles, where the most deprived areas are in

[3] See www.officeforstudents.org.uk/data-and-analysis/institutional-performance-measures/technical-documentation/

[4] The DfE does not accept responsibility for any inferences or conclusions derived from the NPD data by third parties

[5] See www.gov.uk/government/statistics/english-indices-of-deprivation-2019

[6] See www.gov.uk/government/statistics/english-indices-of-deprivation-2019

quintile 1 and the least deprived are in quintile 5. For pupils or students domiciled outside of England, the value is set to 'Unknown or N/A', and data is not presented in the interactive tools.

Local or distance learner

12. Local learners are identified by comparing home travel to work area with study travel to work area, which are calculated from home postcode and study postcode respectively. Local students are those whose home address is in the same travel to work area as their provider. Distance learners are those who are not in attendance at the provider for the vast majority of their course. That is, they are studying at a distance from their provider. These definitions are based on IPSTUDYLOCTYPE described in our 2021 Core Algorithms document.[7] This characteristic is used in both continuation measures.

NS-SEC

13. NS-SEC classifies the socio-economic background of students.[8] For those students who are under 21 at the start of their course, NS-SEC is based on the occupation of their highest earning parent. For those students aged 21 or over at the start of their course, NS-SEC is based on the occupation of the student themselves prior to higher education. Occupations are coded using the Standard Occupation Classification (SOC) and then grouped further into these categories:[9]

    a. Higher managerial, administrative and professional occupations

    b. Intermediate occupations

    c. Routine and manual occupations

    d. Never worked and long-term unemployed.

14. It is only collected in the HESA student record and is only used in the FT continuation measure.

Parental higher education

15. Parental higher education is only collected in the HESA student record.[10] It gives the student's response to the question 'Do any of your parents have any higher education

[7] See www.officeforstudents.org.uk/data-and-analysis/institutional-performance-measures/technical-documentation/

[8] Data for this characteristic comes from the HESA variable SEC. See www.hesa.ac.uk/collection/c19051/a/sec

[9] See section 7 Classes and collapses at www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010#classes-and-collapses

[10] Data for this characteristic comes from the HESA variable pared. See www.hesa.ac.uk/collection/c18051/a/pared

qualifications, such as a degree, diploma or certificate of higher education?'. HESA defines parents as 'natural parents, adoptive parents, step-parents or guardians who have brought you up.' It is only used in the FT continuation measure.

Sex and gender

16. The NPD record collects information on gender, where the only response possible is either male or female, and this is what is used for the ABCS access measure. In the ABCS FT and PT continuation measures, this data is taken from the ILR and HESA records, where the field is recorded as 'sex', and it is also possible to have the response 'other'. In the case of the FT continuation measure, there are sufficient students with this response to place them in their own group. In the PT continuation measure, to avoid unnecessarily discarding data from the statistical model and having a group that is too small to use, these responses are grouped in with males as their continuation rate is closest to theirs.

School types

17. School types are defined using the variable NFTYPE from the NPD. Our access model only includes schools that the Department for Education defines as 'state-funded mainstream' schools. This includes the following school types:

    a. Academy 16-19 Converter

    b. Academy 16-19 Sponsor Led

    c. Academy – converter

    d. Academy – Sponsor-led

    e. City Technology College

    f. Community School

    g. Foundation School

    h. Free School – 16-19

    i. Free School – Mainstream

    j. Free School – Studio School

    k. Free School – UTC

    l. Further Education Sector Institution

    m. Voluntary Aided School

    n. Voluntary Controlled School.

TUNDRA

18. TUNDRA (tracking underrepresentation by area) is an area-based measure that uses tracking of state-funded mainstream school pupils in England to calculate young participation.[11] It only applies to students under 21 at the start of their course. For those aged 21 or over, or for students domiciled outside of England, these are placed in the 'Unknown or N/A' group. TUNDRA has replaced POLAR4 in the ABCS access and FT continuation measures. It is not used in the PT continuation measure.

[11] See www.officeforstudents.org.uk/data-and-analysis/young-participation-by-area/about-tundra/

Annex C: ABCS Access – statistical model and results

1. The access outcome measures the proportion of 18- or 19-year-olds entering higher education (sometimes referred to as young participation). Data regarding these students is taken from the DfE's National Pupil Database (NPD) from the summer in which they obtained their key stage four (KS4) qualifications – most commonly GCSEs. We have then tracked these students through to the start of higher education, where we can determine whether they are in the higher education records two or three years later at the age of 18 or 19. This will capture any level or mode of undergraduate study.

2. We have taken data for pupils who obtained their KS4 qualifications in state-funded mainstream schools in the summers of 2012, 2013, 2014, 2015 and 2016 (that is, in the academic years 2011-12 to 2015-16) from the NPD. Using KS4 cohorts up to 2015-16 allows us to capture the most recent 18- and 19-year-old entrants into higher education in the academic year 2019-20. We use KS4 cohorts because they give almost complete coverage of all 16-year-olds in England. In addition, combining data from five cohorts allows us to carry out robust analysis, ensuring that there are sufficient students in each of the characteristic groups to allow us to carry out analysis regarding their access behaviour.

3. The model includes data on 2,738,955 school pupils, 1,140,660 of which entered higher education aged 18 or 19.

The model

4. We have used a binary logistic regression model to predict the probability of entering higher education. The model includes the following characteristics: ethnicity, FSM eligibility, gender, IDACI, IMD and TUNDRA. Table C1 shows the categories within each of the characteristics and the number and proportion of the students in the model who are in each of these categories.

Table C1: Characteristics in the ABCS access model

Characteristic	Category	Total number of individuals in the five cohorts	Per cent
Ethnicity	Bangladeshi	38,515	1.4%
	Chinese	10,140	0.4%
	Indian	66,120	2.4%
	Pakistani	91,300	3.3%
	Any other Asian background	38,285	1.4%
	Black - African	79,430	2.9%
	Black Caribbean	37,240	1.4%
	Any Other Black Background	15,000	0.6%
	White - British	2,053,975	75.0%
	White - Irish	9,155	0.3%
	Any other white background	102,240	3.7%
	Gypsy, Roma or Traveller	5,210	0.2%
	White and Asian	22,675	0.8%
	White and black African	11,670	0.4%
	White and black Caribbean	35,365	1.3%
	Any other mixed background	36,750	1.3%
	Any other ethnic group	36,345	1.3%
	Refused or unknown	49,530	1.8%
FSM eligibility	Not eligible for FSM	2,006,940	73.3%
	Eligible for FSM	699,100	25.5%
	Unknown or N/A	32,910	1.2%
IDACI	Quintile 1 (most deprived)	614,205	22.4%
	Quintile 2	561,665	20.5%
	Quintile 3	524,505	19.2%
	Quintile 4	509,910	18.6%
	Quintile 5 (least deprived)	491,310	17.9%
	Unknown or N/A	37,360	1.4%
IMD	Quintile 1 (most deprived)	619,635	22.6%
	Quintile 2	547,880	20.0%
	Quintile 3	513,210	18.7%
	Quintile 4	500,720	18.3%
	Quintile 5 (least deprived)	520,150	19.0%
	Unknown or N/A	37,360	1.4%
Gender	Female	1,354,445	49.5%

	Male	1,384,510	50.6%
TUNDRA	Quintile 1 (least represented)	540,090	19.7%
	Quintile 2	540,075	19.7%
	Quintile 3	539,860	19.7%
	Quintile 4	539,675	19.7%
	Quintile 5 (most represented)	539,665	19.7%
	Unknown or N/A	39,595	1.5%
Total number of individuals		2,738,955	100%

5. Stepwise selection has been used to determine which two-way interactions to include in the model with entry and stay criteria of α=0.05. This has resulted in the inclusion of the following interactions shown in table C2.

Table C2: Interactions in the ABCS access model

Interactions
Gender*FSM eligibility
Gender*Ethnicity
Gender*IMD
Gender*TUNDRA
Ethnicity*FSM eligibility
Ethnicity*IDACI
Ethnicity*IMD
Ethnicity*TUNDRA
IDACI*FSM eligibility
IMD*FSM eligibility
IMD*IDACI
TUNDRA*FSM eligibility
TUNDRA*IDACI
TUNDRA*IMD

6. The model is:

$$logit(\pi_i) = \beta_0 + \tilde{\beta}_1 ethnicity_i + \tilde{\beta}_2 IDACI_i + \tilde{\beta}_3 IMD_i + \tilde{\beta}_4 FSM_i + \tilde{\beta}_5 gender_i + \tilde{\beta}_6 TUNDRA_i + interactions$$

Where i is an individual, π_i is a binary response variable which takes the value of 1 if the individual accessed higher education aged 18 or 19 and 0 otherwise, $\tilde{\beta}$ represents vectors of different sizes and the interactions are as listed above.

Model results

7. The coefficient estimates for each of the factors and for all the two-way interactions included in the final model can be found in the Excel/CSV files at www.officeforstudents.org.uk/publications/associations-between-characteristics-of-students-abcs-2021/

Annex D: ABCS FT continuation – statistical model and results

1. Full-time continuation is measured one year and 14 days after a student starts their studies. Students are defined as continuing if they are continuing with or have completed their studies or have transferred to another higher education provider to continue their studies.[12]

2. We have used continuation data for UK-domiciled undergraduate students who started their courses between the academic years 2014-15 and 2018-19 and were studying full-time at an English provider. We have used data from the ILR and HESA student and student alternative records. Data regarding FSM eligibility has been taken from the NPD. Combining data from five cohorts allows us to carry out robust analysis, ensuring that there are enough students in each of the characteristic groups to allow us to carry out analysis regarding their continuation behaviour.

3. The model includes data on 2,079,440 students, 1,875,375 of which are recorded as continuing.

The model

4. We have used a binary logistic regression model to predict the probability of continuing in higher education. The model includes the following characteristics: age, care experience, disability, ethnicity, FSM eligibility, IDACI, IMD, local or distance learner, NS-SEC, parental education, sex and TUNDRA. Table D1 shows the categories within each of the characteristics and the number and proportion of the students in the model who are in each of these categories.

Table D1: Characteristics in the ABCS FT continuation model

	Category	Total number of students in the five cohorts	Per cent
Age	18 and under	926,510	44.6%
	19	431,090	20.7%
	20	154,720	7.4%
	21-25	256,150	12.3%
	26-50	287,980	13.9%
	51+	22,990	1.1%
Care experience	Care experienced	12,725	0.6%
	Not care experienced	1,410,675	67.8%

[12] For details of how we calculate continuation measures from the data, see the OfS Institutional performance measures: Regulatory indicators, methodology and rebuild instructions at www.officeforstudents.org.uk/data-and-analysis/institutional-performance-measures/technical-documentation/

	Unknown or N/A	656,045	31.6%
Disability	Cognitive or learning difficulties	119,395	5.7%
	Mental health condition	62,205	3.0%
	Multiple or other impairments	44,265	2.1%
	Sensory, medical or physical impairments	44,715	2.2%
	Social or communication impairment	15,335	0.7%
	No disability reported	1,793,520	86.3%
Ethnicity	Asian or Asian British - Chinese	15,305	0.7%
	Asian or Asian British - Indian	74,940	3.6%
	Asian or Asian British - Pakistani	84,450	4.1%
	Asian or Asian British - other	51,305	2.5%
	Asian or Asian British – Bangladeshi	46,385	2.2%
	Black or black British - African	152,345	7.3%
	Black or black British - Caribbean	44,585	2.1%
	Black or black British – other	12,515	0.6%
	Mixed - other	27,280	1.3%
	Mixed - white and Asian	27,805	1.3%
	Mixed - white and black African	13,035	0.6%
	Mixed - white and black Caribbean	27,405	1.3%
	White	1,435,015	69.0%
	Gypsy or Traveller	450	0.0%
	Other ethnic group	39,780	1.9%
	Unknown or N/A	26,850	1.3%
FSM eligibility	Eligible for FSM	222,790	10.7%
	Not eligible for FSM	1,137,375	54.7%
	Unknown or N/A	719,275	34.6%
IDACI	Quintile 1 (most deprived)	407,035	19.6%
	Quintile 2	406,590	19.6%
	Quintile 3	382,725	18.4%
	Quintile 4	383,545	18.4%
	Quintile 5 (least deprived)	427,495	20.6%
	Unknown or N/A	72,050	3.5%
IMD	Quintile 1 (most deprived)	402,665	19.4%
	Quintile 2	410,100	19.7%
	Quintile 3	379,885	18.3%
	Quintile 4	385,485	18.5%
	Quintile 5 (least deprived)	430,090	20.7%
	Unknown or N/A	71,220	3.4%

Local or distance learner	Distance	7,260	0.4%
	Local	617,705	29.7%
	Neither	1,454,480	70.0%
NS-SEC	Higher managerial, administrative and professional occupations	556,030	26.7%
	Intermediate occupations	261,530	12.6%
	Routine and manual occupations	310,450	14.9%
	Never worked and long-term unemployed	6,185	0.3%
	Unknown or N/A	945,245	45.5%
Parental higher education	Higher education qualification held by parent(s)	732,730	35.2%
	Higher education qualification not held by parent(s)	716,400	34.5%
	Unknown or N/A	630,315	30.3%
Sex	Female	1,164,215	56.0%
	Male	914,225	44.0%
	Other	1,005	0.1%
TUNDRA	Quintile 1 (least represented)	166,970	8.0%
	Quintile 2	223,625	10.8%
	Quintile 3	276,220	13.3%
	Quintile 4	341,380	16.4%
	Quintile 5 (most represented)	440,420	21.2%
	Unknown or N/A	630,830	30.3%
Total number of students		2,079,440	100%

5. Stepwise selection has been used to determine which two-way interactions to include in the model with entry and stay criteria of $\alpha=0.05$. This has resulted in the inclusion of the following interactions as shown in table D2.

Table D2: Interactions in the ABCS FT continuation model

Interactions
FSM eligibility*Parental higher education
Age*FSM eligibility
Age*NS-SEC
Age*Care experience
Age*Parental higher education
Care experience*NS-SEC
Ethnicity*NS-SEC

Ethnicity*Age
Ethnicity*Care experience
Disability*FSM eligibility
Disability*Parental higher education
Local or distance learner*FSM eligibility
Local or distance learner*Age
Local or distance learner*Ethnicity
Local or distance learner*Disability
Local or distance learner*Parental higher education
Parental higher education*NS-SEC
IDACI*FSM eligibility
IDACI*Age
IDACI*Care experience
IDACI*Disability
IMD*NS-SEC
IMD*Ethnicity
IMD*Local or distance learner
TUNDRA*Age
TUNDRA*Ethnicity
TUNDRA*Parental higher education
Sex*Ethnicity

6. The model is:

$$logit(\pi_i) = \beta_0 + \tilde{\beta}_1 age_i + \tilde{\beta}_2 \text{care experience}_i + \tilde{\beta}_3 disability_i + \tilde{\beta}_4 ethnicity_i + \tilde{\beta}_5 FSM\ eligibility_i + \tilde{\beta}_6 IDACI_i + \tilde{\beta}_7 IMD_i + \tilde{\beta}_8 local\ or\ distance_i + \tilde{\beta}_9 NS-SEC_i + \tilde{\beta}_{10} parental\ higher\ education_i + \tilde{\beta}_{11} sex_i + \tilde{\beta}_{12} TUNDRA_i + interactions$$

where i is an individual, π_i is a binary response variable which takes the value of 1 if the individual continued into the second year of their course and 0 otherwise, $\tilde{\beta}$ represents vectors of different sizes and the interactions are as listed above.

Model results

7. The coefficient estimates for each of the factors and for all the two-way interactions included in the final model can be found in the Excel/CSV files at www.officeforstudents.org.uk/publications/associations-between-characteristics-of-students-abcs-2021/

Annex E: ABCS PT continuation – statistical model and results

1. Part-time continuation is measured two years and 14 days after a student starts their studies. Students are defined as continuing if they are continuing with or have completed their studies or have transferred to another higher education provider to continue their studies.[13]

2. We have used continuation data for UK-domiciled undergraduate students who started their courses between the academic years 2013-14 and 2017-18 and were studying part-time at an English provider. We have used data from the ILR and HESA student and student alternative records. Combining data from five cohorts allows us to carry out robust analysis, ensuring that there are enough students in each of the characteristic groups to allow us to carry out analysis regarding their continuation behaviour.

3. The model includes data on 437,420 students, 284,815 of which are recorded as continuing.

The model

4. We have used a binary logistic regression model to predict the probability of continuing in higher education. The model includes the following characteristics: age, disability, ethnicity, IDACI, IMD, local or distance learner and sex. Table E1 shows the categories within each of the characteristics and the number and proportion of the students in the model who are in each of these categories.

Table E1: Characteristics in the ABCS PT continuation model

Characteristic	Category	Total number of students in the five cohorts	Per cent
Age group	18 and under	18,545	4.2%
	19	17,845	4.1%
	20	19,785	4.5%
	21-25	100,010	22.9%
	26-50	251,850	57.6%
	51+	29,285	6.7%
	Unknown or N/A	100	0.0%
Disability	Cognitive or learning difficulties	15,715	3.6%
	Mental health condition	9,500	2.2%
	Multiple or other impairments	15,560	3.6%

[13] For details of how we calculate continuation measures from the data, see the OfS Institutional performance measures: Regulatory indicators, methodology and rebuild instructions at www.officeforstudents.org.uk/data-and-analysis/institutional-performance-measures/technical-documentation/

	Sensory, medical or physical impairments	10,085	2.3%
	Social or communication impairment	1,350	0.3%
	No disability reported	385,215	88.1%
Ethnicity	Asian or Asian British - Chinese	1,315	0.3%
	Asian or Asian British - Indian	8,500	1.9%
	Asian or Asian British - Pakistani	6,175	1.4%
	Asian or Asian British - other	6,885	1.6%
	Asian or Asian British – Bangladeshi	2,200	0.5%
	Black or black British - African	16,065	3.7%
	Black or black British - Caribbean	8,175	1.9%
	Black or black British – other	1,775	0.4%
	Mixed - other	3,655	0.8%
	Mixed - white and Asian	2,110	0.5%
	Mixed - white and black African	1,405	0.3%
	Mixed - white and black Caribbean	3,830	0.9%
	White	361,300	82.6%
	Gypsy or Traveller	85	0.0%
	Other ethnic group	4,620	1.1%
	Unknown or N/A	9,325	2.1%
IDACI	Quintile 1 (most deprived)	75,445	17.3%
	Quintile 2	82,895	19.0%
	Quintile 3	82,410	18.8%
	Quintile 4	78,650	18.0%
	Quintile 5 (least deprived)	73,185	16.7%
	Unknown or N/A	44,840	10.3%
IMD	Quintile 1 (most deprived)	73,375	16.8%
	Quintile 2	83,365	19.1%
	Quintile 3	82,465	18.9%
	Quintile 4	79,990	18.3%
	Quintile 5 (least deprived)	73,840	16.9%
	Unknown or N/A	44,385	10.2%
Local or distance learner	Distance	201,895	46.2%
	Local	111,475	25.5%
	Neither	124,055	28.4%
Sex	Female	255,685	58.5%
	Male	181,740	41.6%

Total number of students		437,420	100%

5. Stepwise selection has been used to determine which two-way interactions to include in the model with entry and stay criteria of α=0.05. This has resulted in the inclusion of the following interactions shown in table E2.

Table E2: Interactions in the ABCS PT continuation model

Interactions
Ethnicity*Age
Ethnicity*Disability
Ethnicity*Local or distance learner
Ethnicity*IDACI
Disability*Age
Disability*Local or distance learner
Local or distance learner*Age
IDACI*Age
IDACI*Local or distance learner
IMD*Age
IMD*Local or distance learner
IMD*IDACI
Sex*Age
Sex*Ethnicity
Sex*Disability
Sex*Local or distance learner
Sex*IDACI
Sex*IMD

6. The model is:

$$logit(\pi_i) = \beta_0 + \tilde{\beta}_1 age_i + \tilde{\beta}_2 disability_i + \tilde{\beta}_3 ethnicity_i + \tilde{\beta}_4 IDACI_i + \tilde{\beta}_5 IMD_i + \tilde{\beta}_6 local\ or\ distance_i + \tilde{\beta}_7 sex_i + interactions$$

where i is an individual, π_i is a binary response variable which takes the value of 1 if the individual continued into the second year of their course and 0 otherwise, $\tilde{\beta}$ represents vectors of different sizes and the interactions are as listed above.

Model results

7. The coefficient estimates for each of the factors and for all the two-way interactions included in the final model can be found in the Excel/CSV files at

www.officeforstudents.org.uk/publications/associations-between-characteristics-of-students-abcs-2021/